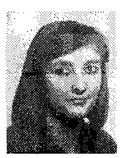
Tijdschrift voor Economie en Management
Vol. XLVII, 4, 2002

How to Design Highly Informative and Cost-Efficient Industrial Experiments

by P. GOOS and M. VANDEBROEK

Peter Goos
Departement Toegepaste
Economische Wetenschappen,
KULeuven, Leuven



Martina Vandebroek
Departement Toegepaste
Economische Wetenschappen,
KULeuven, Leuven

ABSTRACT

Quality has to be built into products and processes during the design stage of their development. It is therefore crucial that any decision made in this early stage of development is based on accurate information. The information is typically obtained from a series of experimental tests, in which the effect of one or more factors on the quality of the product or process is investigated. The settings used for each of the experimental factors at every test run are crucial for the reliability of the information provided by the experiment. In the paper, it is shown how to compute the best factor settings for an industrial experiment. The split-plot type of experiment considered in the paper not only has the advantage that it is easy to carry out, but it has excellent statistical properties as well.

I. INTRODUCTION

The decisions made in the design stage of a new product or a new process have far-reaching consequences for their quality and performance. Making the right decisions in the design stage is therefore crucial for the future success of new products and processes. In industrial environments, these decisions are often based on experiments, the purpose of which is to find the critical factors that influence the performance or the quality of products and processes and to estimate their impact. Industrial experiments typically consist of a small number of runs, in which the quality of the product or process is measured for specific settings of one or more experimental factors. The small number of runs is dictated by the resources available for the experiment, for instance financial or human resources, and time. The design of the experiment consists of determining the factor settings at each individual run.

Once the experiment has been carried out and analysed, the experimenter is able to identify the factors that have a significant impact on the quality of the new product or process and to determine the best settings for each of those factors. Now, the reliability of the analysis of the experiment and the inference that can be made strongly depend on the factor settings that were used in the experiment. In view of the utmost importance of the design stage of new products and processes, it is indispensable that the conclusions drawn from the experiment are as reliable as possible. Finding the factor settings that provide the experimenter with the most reliable information on the factors involved in the experiment is the purpose of the optimal design of experiments.

The purpose of this paper is to show how the approach of optimal design of experiments can be used to design industrial experiments that are both statistically efficient and cost-efficient. In the first part of the paper, we will introduce the concept of a split-plot experiment, which is the type of experiment considered in the paper. Next, we describe an algorithm to design the best possible experiment in two practical instances. Finally, it is shown that the resulting experiments are not only cheaper to conduct, but that they are also statistically more efficient than more expensive experiments.

II. SPLIT-PLOT EXPERIMENTS

In the traditional literature on experimental design, it is recommended that all the runs of the experiment should be conducted in a random fashion and that the factors should be reset independently for each run. The purpose of this randomisation procedure is to avoid that systematic effects like learning effects or time trend effects influence the results obtained from the experiment. Complete randomisation is however time-consuming and expensive, especially when it is difficult or costly to change the settings of some of the experimental factors. Instead, practitioners often arrange the experimental runs so that the settings of the hard-to-change factors are modified as little as possible. One typical example of a hard-to-change factor is the temperature of a furnace because increasing or decreasing furnace temperature is time-consuming and thus costly. Other examples of hard-to-change factors are given in the following three practical instances.

Example 1 The protein experiment

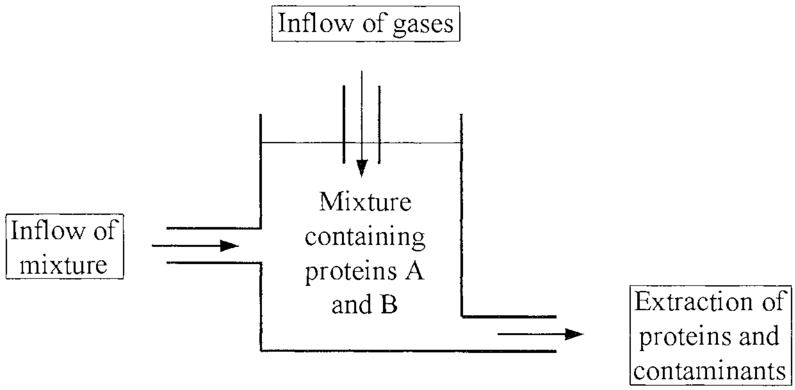
Trinca and Gilmour (2001) describe an experiment to investigate the effect of 5 factors on the yield of a protein extraction process. The factors were the feed position for the inflow of the mixture, the feed flow rate, the gas flow rate, the concentration of protein A and the concentration of protein B. A schematic representation of the extraction process is given in Figure 1. Three levels were used for each factor. Since setting the feed position involved taking apart and reassembling the equipment, it was decided that the feed position should only be changed after one day of experimentation. Thanks to this policy, two experimental runs instead of one could be performed on a single day. Therefore, 42 experimental runs could be carried out in 21 days.

Example 2 The prototype experiment

Bisgaard and Steinberg (1997) use an example from Taguchi (1989) to illustrate how prototype experiments are designed. The purpose of the experiment was to reduce the CO content of exhaust gas. Seven hard-to-change factors, A, B, C, D, E, F and G, each possessing two levels, were studied, along with three driving modes R_1, R_2 and R_3. Due to cost considerations, only 8 of the $2^7 = 128$ combinations of the hard-to-change factor level combinations were used in the experiment. The three driving modes correspond to increasing numbers of rotations per minute. Completely randomising the entire experiment

FIGURE 1

The protein extraction process



was impossible because this would imply that $8 \times 3 = 24$ prototype engines would have to be built, that is one for each experimental run. However, only eight prototype engines were developed in order to save costs and each prototype was used under the three driving modes.

Example 3 The printing ink experiment

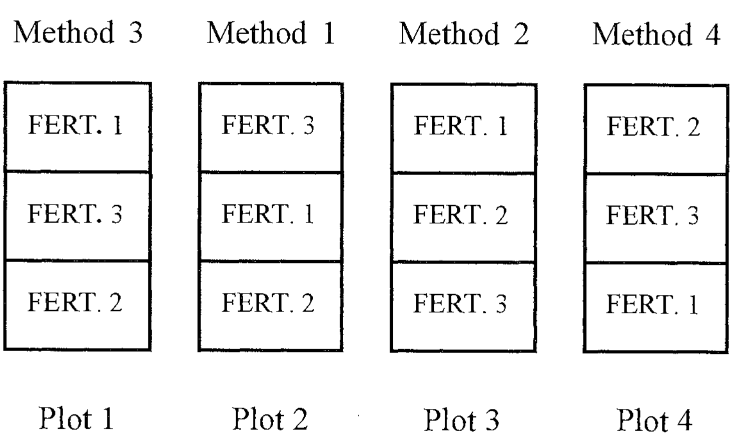
Box and Draper (1987) describe an experiment in which a printing machine's ability to print colouring inks on package labels is examined. The three factors under investigation are machine speed, pressure and distance between the printing tool and the label. The experiment should be conducted in an economical fashion. For this reason, it was decided that the machine speed should be reset as little as possible because modifying the machine speed entails a large set-up time in comparison with the factors pressure and distance.

It is clear that only one hard-to-change factor, namely the feed position of the inflow, is involved in the protein experiment. In the prototype experiment, however, there are 7 hard-to-change factors. In the printing ink experiment, the factor machine speed is the only hard-to-change factor. The common feature of the hard-to-change factors is that it is too cumbersome to reset them independently for each run. In the protein experiment, the feed position is set only once for the 2 runs performed on one day. In the prototype experiment, each prototype is used for several experimental runs.

The experiments obtained by not resetting the levels of the experimental factors are referred to as split-plot experiments. The name split-plot experiment was first used for agricultural experiments in which a full randomisation was impossible. Consider for example an experiment with 12 test runs for investigating the impact of 4 irrigation methods and 3 types of fertiliser on the yield of a crop. Four fields or plots were available. Of course, it was impractical to partition each plot into four parts and to apply a different irrigation method to each part. Therefore, a different irrigation method was randomly applied to each of the whole plots. The 4 plots were then split into 3 so-called sub-plots, to which the fertilisers were assigned at random. The resulting design is displayed in Figure 2. For obvious reasons, the irrigation method is called the whole plot factor of the experiment, whereas the type of fertiliser is called the sub-plot factor. The set of runs conducted on one plot is referred to as a whole plot. Note that all the runs within a whole plot possess the same setting of the whole plot factor.

The traditional agricultural split-plot experiment can easily be translated into an industrial split-plot experiment. As an illustration, consider the prototype experiment of Example 2. With each (whole) plot of the agricultural experiment corresponds a prototype built for a particular combination of the hard-to-change factors A to G. Each prototype is then tested under three driving modes: R_1, R_2 and R_3.

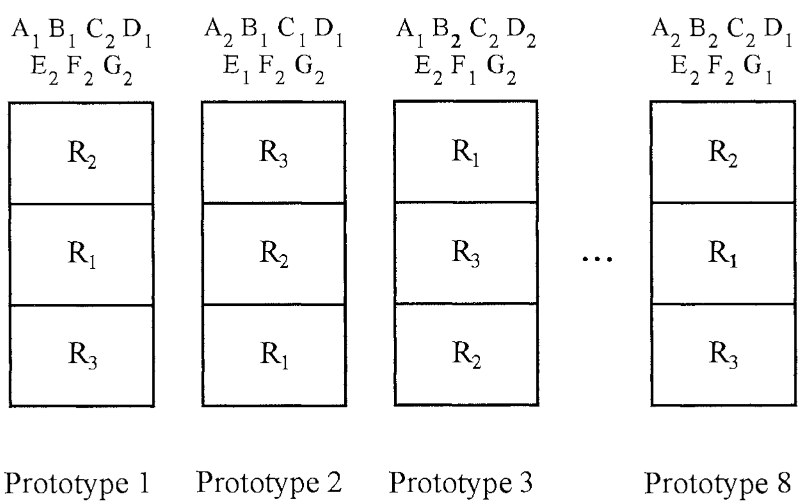
FIGURE 2
An agricultural split-plot experiment



The correspondence between the traditional split-plot experiment and the prototype experiment is graphically displayed in Figure 3. It is clear that the factors A to G will be called the whole plot factors of the prototype experiment and that the driving mode will be referred to as the sub-plot factor.

The protein experiment, the prototype experiment and the printing ink experiment illustrate how the practice of split-plotting considerably simplifies the execution of an experiment. As a consequence, conducting a split-plot experiment is cheaper than a completely randomised experiment and it may save the experimenter substantial amounts of time and money. Therefore, split-plot experiments are heavily used in industry. Nevertheless, no decent tools have been developed to help experimenters in setting up a good split-plot experiment. In order to overcome this problem, experimenters have two options. The first option is to rely on design construction algorithms for completely randomised experiments, e.g. the algorithm of Fedorov (1972) or the algorithm of Atkinson and Donev (1989). These algorithms provide the experimenter with a set of factor settings. These factor settings can be arranged in a sequence so that all the runs with the same level for the hard-to-change factors are grouped. If the experimental runs are conducted in that sequence and the levels of the hard-to-change factors are not reset, the resulting experiment is a split-plot

FIGURE 3
The prototype experiment



experiment. The second option available to the practitioner is to use standard response surface designs in a similar fashion. A brief description of the most popular standard response surface designs, as for instance the central composite designs and the factorial designs, is given in Khuri and Cornell (1987). Standard response surface designs provide the factor settings for completely randomised experiments with a specific number of runs. For experiments with different numbers of runs and for cases with constrained factor settings, researchers often modify these standard design in an arbitrary fashion in order to obtain a feasible design. In the sequel of this paper, it will be shown that this approach can lead to poor designs.

III. DESIGN AND ANALYSIS OF A SPLIT-PLOT EXPERIMENT

In contrast with the observations of a completely randomised experiment, the observations of a split-plot experiment are in most cases not statistically independent. The statistical dependence of the experimental data is of course due to the fact that the factors are not reset independently for each experimental run belonging to the same whole plot. It is explicitly taken into account in the statistical model corresponding to the split-plot experiment. In this section, it is outlined how a split-plot experiment is modelled, how it should be analysed and how the best split-plot design for a particular experimental situation can be computed.

A. *Statistical model*

The observations on the quality or the performance of a product or a process are not statistically independent when they are obtained from a split-plot experiment. Let us denote by y the quality or the performance of the product or process under investigation, by $w_1, w_2, \ldots, w_{m_w}$ or simply by $\mathbf{w}$ the m_w hard-to-change factors presumed to influence the quality or the performance, and by $s_1, s_2, \ldots, s_{m_s}$ or $\mathbf{s}$ the m_s factors that are easier or cheaper to change. The quality or performance, in the sequel of this paper referred to as the response, corresponding to the jth experimental run with the ith level of the hard-to-change factors $\mathbf{w}$ can then be written as

$$y_{ij} = \mathbf{f}'(\mathbf{w}_i, \mathbf{s}_{ij})\boldsymbol{\beta} + \gamma_i + \varepsilon_{ij},\ i = 1, 2, \ldots, b, j = 1, 2, \ldots, k_i,$$

where the function **f** translates the factor settings in first, second or higher order terms, $\boldsymbol{\beta}$ contains the effects of the experimental factors on the response, γ_i is a random whole plot effect corresponding to the ith setting of the whole plot factors, ε_{ij} is a random error reflecting the random variation in each experimental run, b is the number of whole plots in the experiment and k_i is the number of runs within the ith whole plot. It is assumed that the random effects γ_i and ε_{ij} are identically and independently distributed with zero mean and variance σ_γ^2 and σ_ε^2 respectively, and that all the γ_i and ε_{ij} are independent from each other.

Let us now denote by **y** the vector containing the $n = \Sigma k_i$ responses of the experiment, by **X** the so-called design matrix with rows $\mathbf{f'(w}_i, \mathbf{s}_{ij})$, by **Z** a matrix of 0s and 1s assigning the runs to the whole plots, by $\boldsymbol{\gamma}$ vector containing the b whole plot effects and by $\boldsymbol{\varepsilon}$ a vector containing the n random errors. The statistical model can then be written in its matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

The statistical dependence of the data is reflected in the statistical model by the term γ_i, which is common for all the runs within the ith whole plot. The split-plot model assumes that all the runs for which the factors are not reset independently are statistically dependent. In other words, runs belonging to the same whole plot are assumed correlated. Experimental runs obtained by resetting the factors independently remain statistically independent or uncorrelated, as in a completely randomised experiment. Mathematically, the correlation structure of the experimental data can be written as the block diagonal matrix

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_b \end{bmatrix},$$

where each

$$\mathbf{V}_i = \begin{bmatrix} \sigma_\gamma^2 + \sigma_\varepsilon^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\gamma^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\gamma^2 \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 + \sigma_\varepsilon^2 \end{bmatrix} = \sigma_\varepsilon^2 \begin{bmatrix} 1+\eta & \eta & \cdots & \eta \\ \eta & 1+\eta & \cdots & \eta \\ \vdots & \ddots & \ddots & \vdots \\ \eta & \eta & \cdots & 1+\eta \end{bmatrix}$$

represents the correlation structure of the k_i observations in the ith whole plot. The extent to which the observations within one whole plot are correlated is denoted by $\eta = \sigma_\gamma^2/\sigma_\varepsilon^2$. In the sequel of this paper, we will refer to η as the degree of correlation of the split-plot experiment.

B. *Analysis of a split-plot experiment*

The purpose of the experiment is to estimate the effects $\boldsymbol{\beta}$ of the experimental factors on the response of interest. The statistical dependence of the data obtained from a split-plot experiment should be taken into account when estimating the factor effects. This is done explicitly by the generalised least squares estimator

$$\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

which provides better estimates of the factor effects than the ordinary least squares estimator

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

which is appropriate for analysing an experiment with uncorrelated observations. In order to judge whether the estimated factor effects are significantly different from zero, a t-test is used to compare their magnitude to their variance. The variances of the estimated factor effects are given on the diagonal of the variance-covariance matrix

$$\mathrm{Cov}(\hat{\beta}_{GLS}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

The statistical analysis outlined in this section can be performed by many statistical packages commercially available. For instance, the appropriate procedure in the popular SAS package is proc mixed. Further details on the analysis of split-plot experiments can be found in Letsinger, Myers and Lentner (1996) and Goos (2002).

Box and Jones (1992) and Davison (1995) point out that improperly analysing a split-plot experiment as a fully randomised experiment, that is by using ordinary least squares, can lead to erroneously considering significant factor effects as insignificant and vice versa. In other words, an improper analysis of a split-plot experiment may lead to a wrong selection of the critical factors and, hence, to a poorly designed product or process.

C. *Design of a split-plot experiment*

The purpose of an experiment is to estimate the effects of the experimental factors on the performance or the quality of the product or process under investigation. In order to obtain reliable estimates, the experimenter should not only analyse the experiment in a proper way, but he or she should also design it properly. Important information on the quality of the parameter estimates $\hat{\beta}_{GLS}$ obtained from the experiment is provided by the variance-covariance matrix $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$, which contains the variances and the correlation structure of the estimated factor effects. Ideally, the variances of the estimated parameter effects are small and the estimated parameter effects are uncorrelated. In that case, the experiment allows the experimenter to draw reliable conclusions about the significant factors.

The variances and the correlation structure of the estimated parameter effects can be controlled by the experimenter through the design matrix $\mathbf{X}$, which contains the settings of the experimental factors at each run of the experiment. The optimal factor settings are those settings that produce the best properties of the variance-covariance matrix $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. One way to quantify the properties of this matrix is to compute its determinant, which is a measure of the so-called generalised variance of the parameter estimates. Therefore, a good split-plot experiment minimises the determinant of $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ over all feasible factor settings. The resulting experiment is called D-optimal. It should be stressed at this point that the D-optimal designs depend on the degree of correlation η through $\mathbf{V}$. In the next section, an algorithm is described for the computation of D-optimal experiments.

IV. COMPUTING OPTIMAL SPLIT-PLOT EXPERIMENTS

The computation of the best possible split-plot experiment or the optimal design of the experiment is not a simple issue. It is clear that the optimal design must take into account practical considerations as, for instance, the number of experimental runs available, the number of runs that can be performed on one day and the factor settings that are feasible. The number of runs available is usually restricted by a budget or a time constraint. In this section, we discuss how to use a design construction algorithm for computing the best possible design for the experimental situation at hand. We will use the printing ink study from

Example 3 as a practical illustration. In this experiment, the factor machine speed serves as the only whole plot factor of the experiment, whereas pressure and distance are the sub-plot factors. The production time assigned to the experiment allowed for 27 runs and the factor speed could be reset as little as possible. The interest was in estimating the linear and quadratic effects of the experimental factors and the 2-factor interaction effects. The mathematical model under investigation can be written as

$$\mathbf{f}'(\mathbf{w}, \mathbf{s})\boldsymbol{\beta} = \beta_1 + \beta_2 w + \beta_3 s_1 + \beta_4 s_2 + \beta_5 w^2 + \beta_6 s_1^2 + \beta_7 s_2^2 + \beta_8 w s_1 + \beta_9 w s_2 + \beta_{10} s_1 s_2,$$

where w represents the machine speed, s_1 represents the pressure and s_2 represents the distance between the printing tool and the label.

In the sequel of this section, the input for the design construction algorithm will be discussed rather than the technical details. A technical description of the algorithm can be found in Goos and Vandebroek (2001) or in Goos (2002). A FORTRAN 77 implementation and a set of sample input and output files are available from the authors.

The input to the design construction algorithm consists of a detailed description of the experimental situation. The parameters that have to be specified are

- the number of experimental runs n,
- the number of whole plot factors m_w,
- the number of sub-plot factors m_s,
- the form of the mathematical model under investigation,
- the expected degree of correlation.

Typically, information on the expected degree of correlation can be obtained from data obtained from former experimentation of a similar kind. It turns out that degrees of correlation exceeding unity are no exception in split-plot experiments. According to Bisgaard and Steinberg (1997), the whole plot error variance σ_γ^2 is usually larger than the sub-plot error variance a σ_ε^2, so that in many cases $\eta \geq 1$.

In some practical instances, the number of whole plots b and the number of experimental runs k_i within each whole plot is dictated by the experimental situation. In that case, these parameters have to be specified with the input to the design construction algorithm.

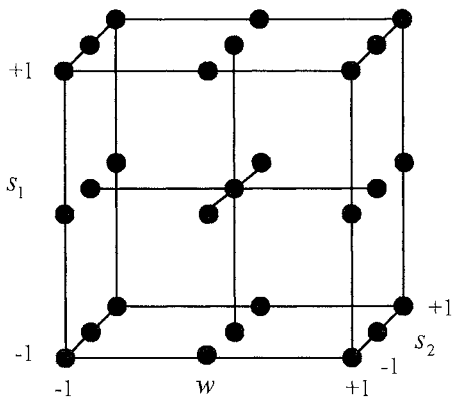
For instance, the number of whole plots in the protein experiment from Example 1 is 21 and the number of runs in each whole plot equals 2. In the printing ink experiment from Example 4, it was only required that the machine speed should be reset as little as possible.

Finally, the experimenter has to specify the possible settings of the experimental factors. More specifically, he or she has to list all possible combinations of settings for all the factors involved. This is particularly important when constraints are imposed on the factor settings and when qualitative factors are involved in the experiment. In other instances, the experimenter can use the default list of candidates. The default list of 27 candidates for the printing ink example is given in Table 1. Three levels are used for the experimental factors because a quadratic term is included in the model for each of them. If no quadratic terms were included in the model, only two levels would have been needed. If cubic terms were included, four levels would have been needed. By default, equidistant factor settings are used. It is usual to denote the factor settings in coded form. In Table 1, the low level of a factor is represented by −1, a 0 represents the middle level of a factor and a +1 represents the high level. Note that the equality of the number of runs and the number of candidates in this example is a coincidence. The 27 candidate combinations of factor settings are graphically displayed in Figure 4. Each combination of settings for w, s_1 and s_2 is a candidate point. The runs of the experiment will be selected from the list of candidates. A selected candidate is referred to

TABLE 1
Coded list of the candidate factor settings for the printing ink example

	w	s_1	s_2		w	s_1	s_2		w	s_1	s_2
1	−1	−1	−1	10	0	−1	−1	19	+1	−1	−1
2	−1	−1	0	11	0	−1	0	20	+1	−1	0
3	−1	−1	+1	12	0	−1	+1	21	+1	−1	+1
4	−1	0	−1	13	0	0	−1	22	+1	0	−1
5	−1	0	0	14	0	0	0	23	+1	0	0
6	−1	0	+1	15	0	0	+1	24	+1	0	+1
7	−1	+1	−1	16	0	+1	−1	25	+1	+1	−1
8	−1	+1	0	17	0	+1	0	26	+1	+1	0
9	−1	+1	+1	18	0	+1	+1	27	+1	+1	+1

FIGURE 4
Graphical representation of the candidate factor settings for the printing ink example



as a design point. Each candidate point can be selected more than once.

V. OPTIMAL DESIGNS

In this section, the features of the optimal split-plot designs for the printing ink example from Example 3 and for the protein example from Example 1 will be discussed. It turns out that these features are valid in other experimental situations as well. Next, the statistical efficiency of the optimal split-plot experiments, which are relatively easy and inexpensive to carry out, will be compared to that of the best completely randomised experiment, which is more costly and more time-consuming to conduct because a properly conducted completely randomised experiment requires independently resetting the factor levels at each run. This can be done by setting the factors to a neutral level between two consecutive runs. Very often, however, factor levels that are identical in two consecutive runs are not reset. It is therefore likely that these two runs are correlated. The resulting experiment is a so-called improperly conducted completely randomised design. Often, runs with the same level for the hard-to-change factors are grouped, such that the hard-to-change factors are reset only on a small number of occasions.

A. *Features of the optimal designs*

As already pointed out in Section 3, the optimal split-plot designs depend on the degree of correlation. For the printing ink example, we have obtained four different optimal designs. The 4 designs are denoted by SPD1 to SPD4 and displayed in Figure 5. SPD1 is the best design option when $0 \leq \eta \leq 0.3959$. SPD2 is optimal when $0.3959 \leq \eta \leq 0.4727$. Finally, SPD3 and SPD4 are the best possible split-plot experiments when $0.4727 \leq \eta \leq 5.7306$ and $\eta \geq 5.7306$ respectively.

The printing ink example demonstrates that precise knowledge of the degree of correlation η is not needed for the construction of optimal split-plot experiments because each design from Figure 5 is optimal for an interval of η-values. The width of the intervals is largest when η exceeds unity. Since, in practice, η is expected to be larger than one, this is extremely important from a practitioner's point of view.

A closer examination of the designs in Figure 5 shows two clear evolutions in the designs. Firstly, the number of runs at the middle level of whole plot factor w decreases with η. It can be verified from the figure that SPD1 possesses 5 runs at $w = 0$, whereas SPD2, SPD3 and SPD4 have only 4, 3 and 2 runs respectively at $w = 0$. Secondly, SPD1 gives the optimal factor settings for an experiment with uncorrelated observations, that is an experiment with $\eta = 0$. In contrast, the factor settings given by SPD2, SPD3 and SPD4 are inefficient for an experiment with uncorrelated data. In other words, the optimal factor settings for a split-plot experiment with correlated data are different from those that are optimal for a properly conducted completely randomised experiment. As a result, taking into account the split-plot structure yields different factor settings.

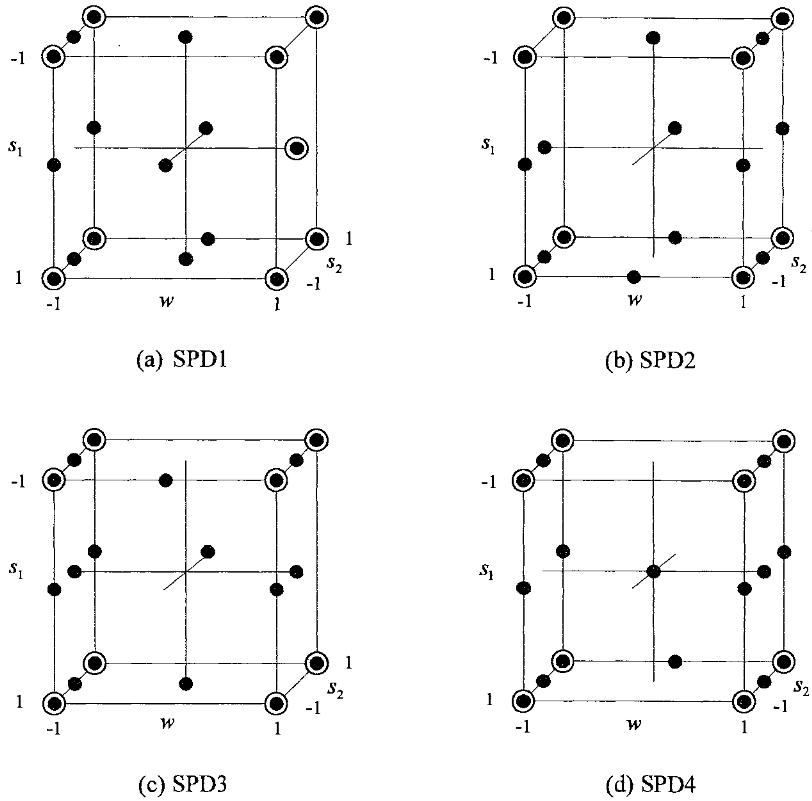
B. *Statistical comparisons*

As already pointed out in the introduction, split-plot experiments are heavily used in industry. Since no tools existed for their efficient design, practitioners had two options. One was to use an algorithm for the computation of an optimal completely randomised experiment.

Another was to use a modified standard response surface design. In this section, the statistical efficiency of the optimal split-plot designs in Figure 5 will be compared to each of these two options. In addition, the optimal design for the protein experiment of Example 1 is

FIGURE 5

Optimal split-plot experiments for the printing ink example. A single bullet represents one experimental run with a particular factor setting. A circled bullet represents two experimental runs with identical factor settings



(a) SPD1 (b) SPD2

(c) SPD3 (d) SPD4

compared to a modified standard response surface design in terms of statistical efficiency.

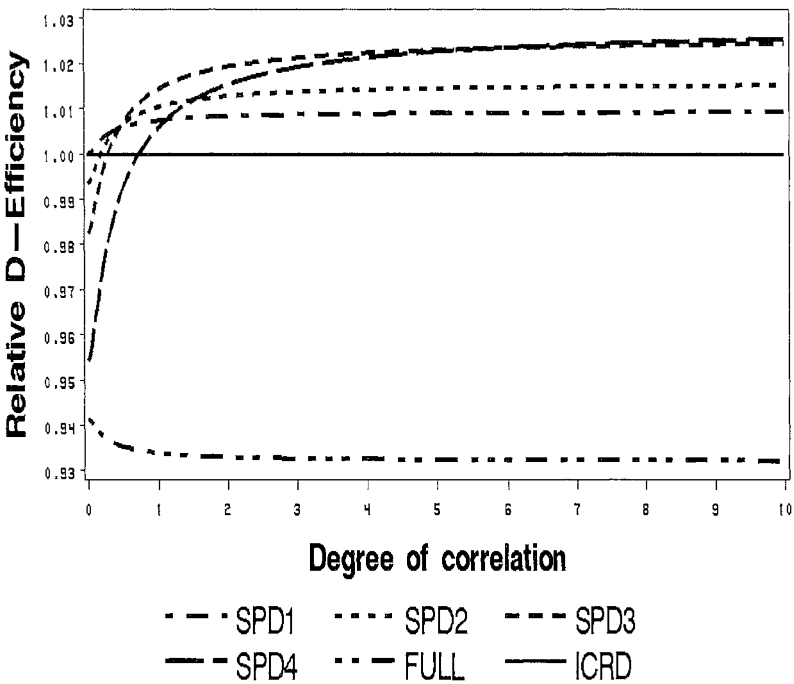
The best possible designs for the printing ink experiment are given by the four designs in Figure 5. One alternative is to use the factor settings produced by an algorithm for computing optimal completely randomised experiments and to conduct the experiment improperly by grouping the runs with the same level of the factor speed and not resetting the speed for all the runs in the same group. The resulting experimental set-up will be referred to as an improperly conducted completely randomised design (ICRD). Another possibility is to use the factor settings given by a standard response surface design, as for instance the 3^3 full factorial design (FULL), in an improper fashion. The design points of the 3^3 design are displayed in Figure 4. Improperly conducting this design leads to three whole plots of nine observations. In Figure 6, the statistical efficiencies of the 6 different design options are compared for η-values between 0 and 10. In the figure, the horizontal benchmark corresponds to the ICRD. It turns out that the 3^3 factorial design is a poor option for any value of η, and that the ICRD is less efficient than the best split-plot design.

A similar comparison can be made for the protein experiment, for which Trinca and Gilmour suggest using the modified central composite design displayed in Table 2 and for which a large η-value was expected. The optimal split-plot design generated by our algorithm is displayed in Table 2 as well. This design is optimal for any practical value of η. It turns out that the modified central composite design is 25% less efficient than the optimal split-plot design. Moreover, the predictions obtained from the optimal split-plot design are 7.1% more precise. These efficiency gains are much larger than in the printing ink example because the benchmark design for this example is poor. This is due to the fact that the benchmark design was constructed by replicating a few points of a standard response surface design.

As a conclusion, the optimal designs perform substantially better in terms of statistical efficiency than the design options currently available for the construction of split-plot experiments. Another efficiency comparison that is extremely important from a practical point of view is that between a split-plot experiment, in which the factors are not reset independently for each run, and an properly conducted completely randomised experiment, in which all factors are reset independently. In Figure 7, the statistical efficiency of the best possible split-plot design for the printing ink experiment (SPD) is compared to a properly conducted completely randomised experiment (PCRD). It is clear that the split-plot design performs much better than the completely randomised experiment as soon as η exceeds one. As a result, the split-plot design is not only easier and cheaper to conduct but it is also statistically more efficient than the completely randomised experiment. The split-plot design is twice as efficient when η is about 6 and thrice as efficient when $\eta = 10$.

FIGURE 6

Comparison of the efficiency of the optimal split-plot designs from Figure 5 with a design obtained by ignoring the correlation structure (ICRD) and a 3^3 full factorial design (FULL)



VI. FURTHER IMPROVEMENTS

In the previous section, it was already shown that split-plot experiments are often statistically more efficient than completely randomised experiments. It turns out that the statistical efficiency of the split-plot experiments can be further improved by slightly increasing the number of times the hard-to-change or whole plot factors are reset independently. The drawback of this strategy is of course that the cost of the split-plot experiment will increase.

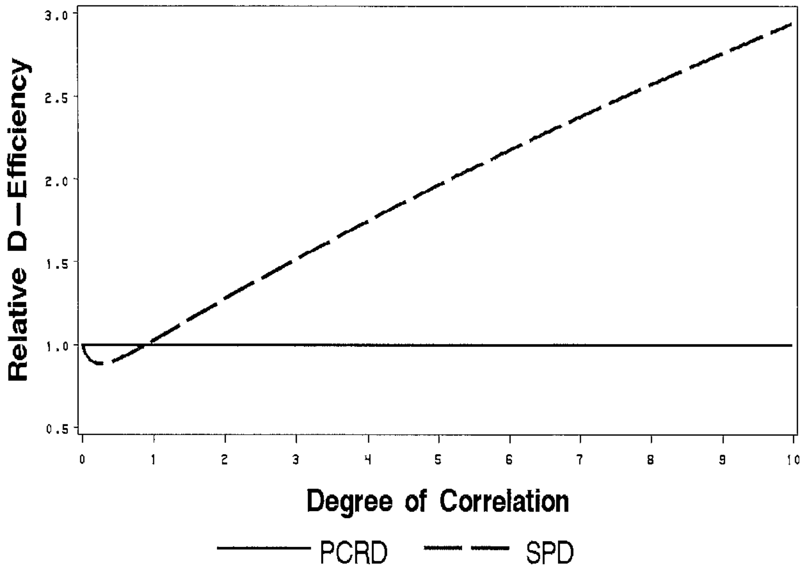
Another important advantage of resetting the whole plot factors more often is that the experimenter is better protected against failures of the experiment. When something goes wrong at a certain setting of the whole plot factors, all the runs within the corresponding whole plot are affected and a substantial amount of information may be lost.

TABLE 2
Design options for the protein experiment

Whole Plot	Central Composite Design					Optimal Split-plot Design				
	w	s_1	s_2	s_3	s_4	w	s_1	s_2	s_3	s_4
1	−1	−1	−1	−1	−1	−1	−1	−1	−1	+1
	−1	+1	−1	−1	+1	−1	+1	−1	+1	−1
2	−1	−1	+1	−1	+1	−1	0	+1	0	−1
	−1	0	0	0	0	−1	−1	−1	−1	−1
3	−1	−1	0	0	0	−1	−1	+1	−1	+1
	−1	+1	+1	−1	−1	−1	+1	0	+1	0
4	−1	−1	+1	+1	−1	−1	+1	+1	+1	−1
	−1	+1	+1	−1	+1	−1	0	0	−1	0
5	−1	0	0	0	−1	−1	−1	+1	+1	−1
	−1	0	+1	0	0	−1	+1	−1	−1	−1
6	−1	0	0	0	0	−1	0	−1	0	−1
	−1	+1	−1	+1	−1	−1	+1	+1	−1	+1
7	−1	0	−1	0	0	−1	−1	−1	+1	−1
	−1	+1	+1	+1	+1	−1	+1	+1	+1	+1
8	0	−1	−1	−1	+1	−1	−1	−1	+1	+1
	0	+1	−1	+1	−1	−1	−1	+1	−1	−1
9	0	−1	−1	+1	−1	−1	+1	−1	−1	+1
	0	+1	−1	+1	+1	−1	−1	+1	+1	+1
10	0	−1	−1	+1	+1	−1	+1	0	−1	−1
	0	+1	−1	−1	−1	−1	+1	−1	+1	+1
11	0	−1	+1	−1	−1	0	−1	0	0	+1
	0	+1	+1	+1	+1	0	0	−1	+1	0
12	0	−1	+1	+1	+1	0	+1	+1	0	0
	0	+1	+1	+1	−1	0	+1	−1	−1	+1
13	0	−1	0	0	0	0	+1	−1	0	0
	0	0	0	0	+1	0	0	+1	−1	−1
14	0	0	0	+1	0	+1	−1	−1	0	−1
	0	+1	0	0	0	+1	−1	+1	−1	+1
15	+1	−1	−1	−1	−1	+1	−1	−1	−1	+1
	+1	+1	0	0	0	+1	+1	+1	−1	−1
16	+1	−1	−1	+1	+1	+1	−1	−1	+1	+1
	+1	0	+1	0	0	+1	+1	+1	+1	−1
17	+1	−1	+1	+1	−1	+1	0	+1	+1	+1
	+1	0	0	−1	0	+1	−1	+1	−1	−1
18	+1	−1	+1	−1	+1	+1	−1	+1	+1	0
	+1	+1	+1	−1	−1	+1	+1	0	−1	+1
19	+1	0	0	−1	0	+1	−1	0	+1	−1
	+1	0	−1	0	0	+1	+1	−1	−1	0
20	+1	+1	−1	−1	+1	+1	+1	−1	+1	+1
	+1	0	0	0	−1	+1	0	−1	−1	−1
21	+1	0	0	0	+1	+1	+1	+1	0	+1
	+1	0	0	+1	0	+1	+1	−1	+1	−1

FIGURE 7

Comparison of the statistical efficiency of an optimal split-plot design (SPD) with a properly conducted completely randomised design (PCRD)



Resetting the whole plot factors more often leads to smaller whole plots and therefore to a smaller loss of information when problems occur at a given whole plot factor setting.

VII. DISCUSSION

In this paper, an approach was presented for the computation of the best possible split-plot experiment for a particular industrial situation. Split-plot experiments have become very popular in industry because they do not require that the factor levels are reset independently for each run. For this reason, split-plot experiments are easier to conduct than completely randomised experiments, in which the factor levels are reset independently for every run of the experiment. An algorithmic approach for the optimal design of split-plot experiments was described and the resulting designs were examined. It was shown that substantial efficiency gains can be achieved over existing design options. In view of the utmost importance of the design stage in the development of new products and processes, this is an important result. Moreover, split-plot experiments are not only easier, and thus less expensive, to conduct than completely randomised experiments, but they are in many practical instances also statistically more efficient.

ACKNOWLEDGEMENT

The research that led to this paper was carried out while the first author was a Postdoctoral Researcher of the Fund for Scientific Research Flanders.

REFERENCES

Atkinson, A.C. and Donev, A.N., 1989, The Construction of Exact D-Optimum Experimental Designs with Application to Blocking Response Surface Designs, *Biometrika* 76, 515-526.

Bisgaard, S. and Steinberg, D.M., 1997, The Design and Analysis of 2^{k-p} *s* Prototype Experiments, *Technometrics* 39, 52-62.

Box, G. and Draper, N.R., 1987, Empirical Model-Building and Response Surfaces, (Wiley, New York).

Box, G.E.P. and Jones, S.P., 1992, Split-Plot Designs for Robust Product Experimentation, *Journal of Applied Statistics* 19, 3-26.

Cornell, J.A. and Gorman, J.W., 1984, Fractional Design Plans for Process Variables in Mixture Experiments, *Journal of Quality Technology* 16, 20-38.

Davison, J.J., 1995, Response Surface Designs and Analysis for Bi-Randomization Error Structures, Ph.D. thesis, (Virginia Polytechnic Institute and State University).

Fedorov, V.V., 1972, Theory of Optimal Experiments, (Academic Press, New York).

Goos, P., 2002, The Optimal Design of Blocked and Split-Plot Experiments, (Springer-Verlag, New York).

Goos, P. and Vandebroek, M., 2001, Optimal split-plot designs, *Journal of Quality Technology* 33, 436-450.

Khuri, A.I. and Cornell, J.A., 1987, Response Surfaces: Designs and Analyses, (Marcel Dekker, New York).

Letsinger, J.D., Myers, R.H. and Lentner, M., 1996, Response Surface Designs for Bi-Randomization Structures, *Journal of Quality Technology* 28, 381-397.

Taguchi, G., 1989, Introduction to Quality Engineering: Designing Quality into Products and Processes, (Asian Productivity Organization, Tokyo).

Trinca, L.A. and Gilmour, S.G., 2001, Multi-stratum Response Surface Designs, *Technometrics* 43, 25-33.